Quantifying Information Leakage in RFID Systems

Xu Huang
School of Information Sciences and Engineering
University of Canberra
ACT 2601, Canberra, Australia
Xu.Huang@canberra.edu.au

Abstract - **Radio Frequency Identification (RFID) systems provide large scale, automated tracking solutions and superior reliability over existing tracking systems as well as the possibility of authentication, but also pose a threat to customer privacy, which already drew great attentions from researchers in this field. In this paper the quantifying information leakage in RFID systems will be first investigated via the Shannon's information theory, and the analysis results are also extended from binary to *r*-ary coding systems. The secondary contribution of this paper is that based on the first part discussion a modified "hash-chain" protocol is presented that decreasing the heavy burden on back-end database to authenticate tags, which the normal hash-chain protocol computes hash function many times on every tag.**

Index Terms - RFID, information leakage, cryptographic approach, hash function.

I. Introduction

Recently radio frequency identification (RFID) attracts great attentions as an alternative to the bar code in the distribution industry, supply chain and banking sector. RFID systems also increasingly are used in control and tracking, medical monitor systems, and other daily managements and businesses. This is because RFID system that has advantages of contact-less type and can hold more data than the bar code. Nevertheless, an automatic identification technology using RFID can suffer from the privacy problems such as tracking without user's recognition [1-2].

The ubiquity of RFID tags also poses many security threats, such as denial of service, tag impersonation, malicious traceability, and information leakage. In particular in everyday life, people are prone to carrying various objects around with them. Some of them are quite personal, which contains the user's privacy. For example, expensive products, cash, medicine information, reading materials, even books that may reflect personal consciousness and avocation. Also there is a function of a RFID system titled as "behavioural tracking and identification", namely an adversary can link the tags with that purchased items and trace the people who have made the purchasing. Avoiding eavesdropping can be done by establishing a secure channel between the tag and the reader. This requires the establishment of a session secret key, which is not always an easy task considering the very limited devices' capacities. The difficulty is supported by the fact that tags and reader do not share a master key in most of the application. As many papers showed that the problem of secure pairing of wireless devices has been tackled by several researchers.

In the [3], the tags that auto-ID centre supports have the following property: each tag has a unique 8-bit password and upon receiving the password, the tag erases itself. That is the fact that the tags in the RFID system suicide preventing any subsequent useful services. Since tags kill command feature and the benefits of RFID are dimmed. The system needs a conscious operating as there is no return back processing.

There is a scheme called "hash lock scheme" as described in [4-7]. The scheme is low-cost because it requires only a hash function. Each tag verifies the reader as follows. The reader has key k for each tag, and each tag holds the result ID. However, ever ID is fixed. So attacker can track the tags and the protocol is vulnerable to reply an attack. Also when attacker disguises the right reader and receives the ID from the tag then disguises the right tag and gets the key from the right reader sending this ID.

Papers [7-8] made an extension of the hash lock type scheme to the so-called "randomized hash lock scheme" that requires the tag to have a hash function and a pseudo-random generator. Each tag calculates the hash function based on the input from pseudo-random generated, r and id, i.e. $c = hash(id|r)$. The tag then sends c and r to the reader. The reader sends the data to the back-end database. Since id_k is sent to the tag through the insecure channel, the tag can be tracked.

There is a scheme titled as "anonymous ID" showed in [9], where the output of a tag is an anonymous ID. The adversary can never know the real ID of the tag because it uses public-key encryption schemes or symmetric-encryption schemes or random value linked to tag's ID on external computation units. However an authentication or secure channel must be established between the reader and the back-end database. Also the anonymous ID is fixed, tracking becomes possible.

The "external re-encryption protocol" [10] seems theoretically more secure than some security protocols because it also uses a public key cryptography technique to protect the tag's ID. It is noted that for the external devices such as a reader to perform encryption and decryption one hand the public key cryptography needs much computation and on the other hand the tags can't compute the public key encryption and decryption. The encrypted tag's ID is fixed in this protocol it has a problem that the tag's data is often rewrite to protect the secret information. Also this protocol is

not suitable for ubiquitous environment as the protocol needs the external devices and the user's action for re-encryption.

Juels, discussed security and privacy [20] and Avoine presented an interesting discussion in security and privacy in RFID systems [21].

There are other approaches [11-15], for example, "noisy tags" [11] where noisy tags are owned by the reader's manager and set out within the reader's field. They are regular RFID tags generate noise on the public channel between the reader and the queried tag, such that an eavesdropper cannot differentiate the messages sent by the queried tag from the ones sent by the noisy tag. Therefore, she will be unable to identify the secret bits that are sent to the reader. The complexes would make it hard to be real life application.

In this paper, we first investigated, via the Shannon's information theory, an analysis result and then it is also extended from a binary system to *r*-ary coding system. This paper is to present a modified titled "hash-chain" protocol that is decreasing the heavy burden on back-end database to authenticate tags, which saves time in comparison with the other hash-chain protocols.

The rest sections will be as follows, in section 2 the quantifying information leakage in RFID systems will be presented via the Shannon's information theory. It will show that Shannon's equations can be extended what we need for the RFID systems. Then we use the Shannon's information theory to explain the quantifying information leakage in this section.

A conclusion will be obtained based on the discussion of Shannon's information theory that when the communication channel is built with pure Gaussian noisy the noise entropy.

In section 3, a modified hash-chain protocol is established based on the cryptographic approach, where the hash-chain protocol will save a heavy burden of calculations on back-end database to authenticate tags due the fact that instead of using bit transiting a coding system is used.

We shall show the conclusion of this paper in section 4.

II. Quantifying information leakage

In order to compare our final results with that obtained from similar methods we shall follow the terminology terms used in those papers, e.g. [17]. Let's consider a memoriless source m emitting messages m_1, m_2, ..., m_n with probabilities P_1, P_2, ..., P_n, respectively $(P_1 + ... + P_n) = 1$. Here, a memoriless source implies that each message emitted is independent of the previous message(s). By this definition we have the information content of message m_i is I_i, given by the following equation:

$$I_i = \log_2 \frac{1}{P_i} \tag{1}$$

The average information per message of a source in m is defined as its entropy, denoted by $H(m)$. Therefore, we have

$$H(m) = \sum_{i=1}^{n} P_i I_i \quad \text{bits} \tag{2}$$

If the channel is noiseless, then the reception of some symbol y_j uniquely determines the message transmitted. Because of noise, there is a certain amount of uncertainty regarding the transmitted symbol when y_j is received. If $P(x_i|y_j)$ represents the conditional probabilities that x_i was transmitted when y_j is received, then there is an uncertainty of $\log_2[1/P(x_i|y_j)]$ about x_i when y_i is received. When this uncertainty is averaged over all x_i and y_j, we obtain $H(x|y)$, which is the average uncertainty about a transmitted symbol when a symbol is received. Therefore, we have

$$H(x \mid y) = \sum_i \sum_j P(x_i, y_j) \log_2 \frac{1}{P(x_i \mid y_j)} \tag{3}$$

Hence, if the channel were noiseless, the uncertainty would be zero. Obviously, this uncertainly, $H(x|y)$, is caused by channel noise, by which we lose an average of $H(x|y)$ bits of information per symbol. Therefore, in the transaction the amount of information the receiver receives is, on the average, $I(x;y)$ bits per received symbol, we have

$$I(x; y) = H(x) - H(x \mid y) \tag{4}$$

Here, $I(x;y)$ is the mutual information of x and y. From the equations (3), for the continuous channel we have [18]:

$$H(y \mid x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) \log_2 \frac{1}{p(y \mid x)} dxdy \tag{5}$$

If $p_n(.)$ represents the PDF of noise sample n, then

$$p(y \mid x) = p_n(y - x) \tag{6}$$

Then, we have

$$\int_{-\infty}^{\infty} p(y \mid x) \log_2 \frac{1}{p(y \mid x)} dy = \int_{-\infty}^{\infty} p_n(y - x) \log_2 \frac{1}{p_n(y - x)} dy$$

Letting y-x = z and note the right-hand side is the entropy $H(n)$ we have

$$H(y \mid x) = H(n) \int_{-\infty}^{\infty} p(x) dx = H(n) \tag{7}$$

Therefore, we have:

$$I(x;y) = H(y) - H(n) \tag{8}$$

It is easy to obtain the maximum entropy for given mean square value of x via "undetermined multipliers" the maximum entropy, (or maximum uncertainty) is when the

distribution of x is Gaussian distribution. Hence, we have the case that if we submit Gaussian distribution entropy we have

$$H(n) = \frac{1}{2}\log_2 2\pi e\sigma^2 = \frac{1}{2}\log_2 17.1\sigma^2 \quad (9)$$

Now let's have a closer look at equation (7) in a RFID system it represents the average information to link one output from an RFID device with the output history by the attacker. In fact it is easily to find that this term is the exactly defined "measure unlinkability" in [16]. In fact we see the that from Shannon's information theory, $H(n)$ is the entropy of noise, or maximum noise between information source and the information receiver. Hence, the bigger of $H(n)$ is, the more unlinkability will be and the information will be more safe. Therefore, we can use the $H(n)$ as a parameter, called the "unlinkability", to measure the safety in a RFID detecting system. It is obvious that if $H(n) = 0$ there is no noise at all hence the "unlinkability" is minimum and the information will be totally obtained by the attacker.

It is noted that if the RFID communication channel is wider, therefore the σ^2 of the Gaussian will be bigger and leads the noise entropy larger to make the information safer as shown by equation (9).

It is important to note that so far what we discussed being in binary coding system. If we use r-ary system, what will happen? In fact, it would be easily to be extended. Because each r-ray digit can assume r values, k r-ray digits can form a maximum of r^k distinct code words. So that to encode each of the n equiprobable messages, we need a minimum of $k = \log_r n$ r-ary digits. Hence the information I per message is

$$I = \log_r \frac{1}{P} \quad r\text{-ary units} \quad (10)$$

From equation (1) and (10) we have

$$1\ r\text{-ary unit} = \log_2 r \text{ bits} \quad (11)$$

Now we can use equation (11) to extended binary cases described above to r-ary systems.

III. A modified hash-chain protocol

In order to check the leakage in RFID cryptographic systems discussed above we are now presenting a modified hash-chain protocol as a case study in this section.

It was presented that the security risks of low-cost RFID tags, the researchers have made a lot of discussions such as Weis *et al.* [4], these papers presented privacy and security risks and how they improved to the setting of low-cost RFID devices in various ways.

In fact, we may use the definition of privacy in [22], where it is the degree to which two authentication sessions of the same tag are not linkable. An authentication session is defined as the interaction between a reader (legitimate or rogue) and a tag at the protocol level. Sessions are defined as "*unlikable*" if an attacker cannot discover whether two responses originated from the same tag with a probability better than random guessing. The *highest degree of unlinkability* exits if any pair of tags is indistinguishable. It is normally to measure privacy as the degree to which a member of the group is indistinguishable from other elements of the group.

Let's have a closer look at a processing that "how to lock tag". Accepting the fact that resource limitations of low-cost tags, people normally offer a simple security scheme based on one-way hash function. As example, to lock a tag, a tag owner stores the hash of a random key as the tag's ID, i.e. ID ← hash(key). This may occur either over the RF channel or a physical contact channel for added security. After locking a tag, the owner stores both the key and an ID in a back-end database. Upon receipt of an ID value, the tag enters its *locked state*. While a tag is locked, the tag responds to all queries with only its ID and offers no other functionality. To unlock a tag, the owner queries the ID from the tag, looks up the appropriate key in the back-end database and finally transmits the key to the tag. The tag hashes the key and compares it to the stored ID. If the values match, it unlocks itself and offers its full functionality to any nearby readers.

It is well known that the hash-lock scheme only requires implementing a hash function on the tag and managing keys on the back-end. This is a relatively low-cost requirement and may be economical in the near future. This scheme may be extended to provide access control for multiple users or to other tag functionality, such as write access. However, we may need fist to check the general design idea of an approach to RFID system.

Let's investigate an approach to protect RFID system from user privacy to which elements in the group are distinguishable and can be measured in bits. In general, the design should meet the following items: (a) keep complete user privacy (b) eliminate the need for extraneous rewrites of the tag information (c) minimize the tag cost (d) eliminate the need for high power of computing units (e) provide forward security.

Following the above descriptions our proposed RFID privacy protection scheme in this paper can be shown as Figure 1, where H and G are one-way hash functions as used in [19]. The reader sends a_i to the back-end database. The back-end database maintains a list of pairs (ID, s_i), where s_i is the initial secret information and is different for each tag. So the back-end database that received tag output a_i from the reader calculates $a'_i = G(H^i(s_i))$ for each s_i in the list and checks if $a_i = a'_i$. if it thje case a'_I such that $a'_i = a_i$, then return the ID, which is a pair of a'_i. The scheme satisfies the security requirements, i.e., indistinguishability and forward security, as follows. G is a one-way function, so if the adversary obtains tag output a_i, one cannot know s_i from a_i. G outputs random values, so if the adversary watches the tag output, the attacker cannot link a_i and a_{i+1}. H is a one way

function, so if the adversary tampers with a tag and obtains the secret information in the tag, the attacker cannot know s_i from s_{i+1}.

From the view point of efficiency, the proposed scheme is efficient enough to yield low-cost RFID tags, since it uses only hash operations that require a small gate size. Therefore, the proposed scheme is reasonable practical for low-cost RFID tags, while still ensuring privacy even in the face of tampering.

We assume that the adversary may eavesdrop on the radio frequency signals between the reader and the tag. The adversary can acquire the secret information stored in the tag to tampering the tag. The RFID scheme should be able to protect the user privacy against such an adversary. To ensure the anonymity of the tag ID, obviously the tag should not output its ID nor should output any constant data. The scheme should meet this requirement. Moreover, this scheme offers the properties of indistinguishability and forward security.

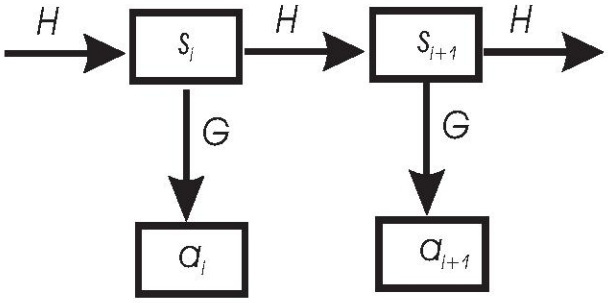

Figure 1: RFID tag sends answer $a_i = G(s_i)$ and renews its secret $s_{i+1} = H(s_i)$

The scheme proposed in Figure 1 is a kind of "hash-chain protocol", which has a heavy burden on back-end database to authenticate tags since the protocol computes hash function i times on every tag. We may find that the Figure 1 can be expressed as Figure 2.

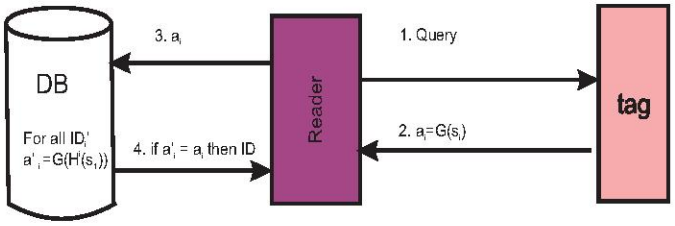

Figure 2: Protocol from Figure 1

In order to modify the Figure 2 and improve this protocol to be scaleable, we have to have a closer look at a RFID working concept.

Privacy concerns are rather moot if someone can remove a tag or steal the item it is attached to without detection. The key point is that tags cannot be trusted to store long-term secrets, such as shared keys, when left in isolation. Tags may also be equipped with a physical contact channel, as found on smart cards, for critical functions or for "imprinting" tags with secret keys.

Additionally, we may assume the tag packaging contains some optical information such as a barcode or human-readable digits. This information may corroborate tag data, as in the design presented in.

Tag readers are assumed to have a secure connection to a back-end database. Although readers may only read tags from within the short, say 3 meters, tag operating range, the reader-to-tag, or forward channel is assumed to be broadcast with a signal strong enough to monitor from long-range, say perhaps 100 meters. The tag-to-reader or backward channel is relatively much weaker, and may only be monitored by eavesdroppers within the tag's shorter operating range. Generally, it will be assume that eavesdroppers may only monitor the forward channel without detection. This relationship is illustrated in Figure3.

It is noted that there are two regions from Figure 3, we can treat those two region differently we propose that in the secure channel we may use "instead of transmitting hash function "bits" transmitting "codes" to make modified protocol be "scaleable".

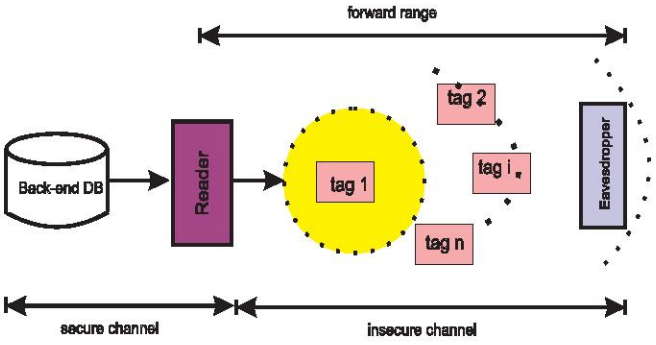

Figure 3: A RFID system

The combination of Figure 2 and Figure 3 suggests that we may take a protocol can be work as that when tag receives a query, the answer $a_i = G(s_i)$ sending to reader, which is the same case that described as that shown in Figure 2. Then reader will code the a_i in terms of H hash function but the output is "coded" information such as CDMD, as an example rather than "bit" information.

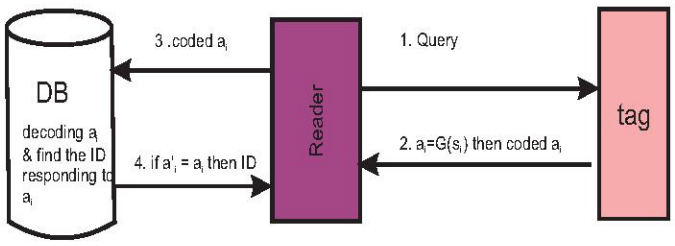

Figure 4: a Modified hash-chain protocol

Then the back-end database, DB, has the list of the coded against ID. So there is no need to calculate all "i" times to check the if a'_i equal to a_i. This modified hash function protocol can be expressed as shown in Figure 4.

For a typical RFID system, the transmission rate in 13.56 MHz and 900 MHz bands available to the RFID tags is approximately 26 kbps / 50 tags at 13.56 MHz and 128 kbps /200 tags at 900 MHz. It was reported that a symmetric encryption algorithm can be constructed with about 6 to 13 kilo-gates [19], so this modified hash function protocol would be, in terms of size of gates, achieved in less than that at the symmetric encryption algorithms.

It is noted that if the hash function has enough length bits that will make the insecure channel close to the Gaussian distribution. Therefore from the noise entropy, equation (9) we have $0.5\log_2 17.1\sigma^2$ bits per message (in the binary system).

IV. Conclusions

We have investigated the cryptographic approach to a RFID tags detecting system. We follow the Shannon's information theory and showed that we can directly establish a parameter called "unlinkability" to measure the safety in a RFID detecting system.

The Shannon's information theory can be used to explain the quantifying information leakage well as shown in section 2.

Based on the discussion of unlinkability, we also have made an extensibility of investigating the Shannon's information theory, by which it is showed when the communication channel is built with pure Gaussian noisy the noise entropy in the communication will make the information leakage minimum and the tag's privacy is at the safest state.

Our paper first discussed the r-ray wireless RFID communication system and extended the results to a r-ray RFID system. In our section three a modified hash-chain protocol is presented based on the cryptographic approach discussed in previous sections, where modified hash-chain protocol will save a heavy burden on back-end database to authenticate tags due the fact that instead of using bit transiting a coding system is used. .

REFERENCES

[1] Associated Press, Benetton undercided on use of "smart tags",8 April 2003.

[2] CNET, Wal-wart cancels "smart shelf" trial, http://www.cnet.com, July. 2003.

[3] Ari Juels and Stephen Weis, Defining strong privacy for RFID, 2006, http://theory.lcs.mit.edu/~sweis/pdfs/JuelsWeis-RFID-Privacy.pdf

[4] Stephen A. Weis, Sanjay E. Sarma, Ronald L. Rivest and Daiel W. W Engels, Security and privacy aspects of low-cast radio frequency identification systems," first international Conference on Security in Pervasive Computing, 2003 http://theory.lcs.mit.edu/sweis/spc-rfid.pdf

[5] S. E. Sarma, S. A. Weis, and D. W. Engels, RFID systems, security and privacy implications, White paper MIT-AUTOID-WH-014, MIT AUTO-ID CENTER, 2002.

[6] S. E. Sarma, S. A. Weis, and D. W. Engels, RFID systyems and security and privacy implications, CHES 2002, LNCS 2523, pp454-469, Springer-Verlag, 2003

[7] S.A. Weis, Security and privacy in radio-frequency identification devices, MS Thesis, MIT, May 2003.

[8] S.A. Weis, S.E. Sarma, R. L. Rivest, and D. W. Engels, Security and privacy aspects of low-cast radio frequency identification systems," Security in Pervasive Computing 2003, LNCS 2802, pp201-212, Springer-Verlag, 2004.

[9] Shingo Kinosita, Fumitaka Hoshino, Tomoyuki Komuro, Akiko Fujimura and Miyako Ohkubo, Nonidentifiable anonymous-ID scheme for RFID privacy protection, CSS 2003 in Japanese.

[10] A. Juels, R. Pappu, Squealing Euros: Privacy protection in RFID-enabled banknotes, Financial cryptography'03, LNCS 2742, pp103-121, Springer-Verlag, 2003.

[11] Claude Castelluccia and Gildas Avoine, Noisy tags: A pretty good key exchange protocol for RFID tags, The 8th International Conference on Smart Card Research and Advanced Applications (*CARDIS*), Tarragona, Spain, April 19-21, 2006, LNCS, Springer-Verlag, 2006.

[12] Sanjay E. Sarma, Stephen A. Weis and Daiel W. Engels, Radio-frequency identification systems, In processing of CHES'02, pp454-469. Springer-Verlag, 2002. LNCS no.2523.

[13] Ari. Juels, Ronald. L. Rivest and Michael Szydlo, The blocker tag: Selective blocking of RFID tags for consumer privacy, In Proceedings of 10th ACM Conference on Computer and Communications Security (CCS 2003), Oct. 2003.

[14] Phillippe Golle, Markus Jakobsson, Ari Juels and P. Syverson, Universal re-encryption for mixnets, 2002, http://www.syverson.org/univrenc-ctrsa.pdf

[15] G. Avoine, Adversarial model for radio frequency identification, 2005. Cryptology ePrint Archive, Report 2005/049. Referenced 2006 at the http://eprint.iacr.org

[16] Yasunobu Nohara, Sozo Inoue, Kensuke Baba, and Hiroto Yasuura, Quantitative evaluation of unilinkable ID matching schemes, In Workshop on Privacy in the Electronic Society-WPES, 2006.

[17] C. E. Shannon, A mathematical theory of communication, 1948.

[18] B.P. Lathi, Modern Digital and Analog Communication Systems, 3/e, by Oxford University Press, Inc. 1998.

[19] Miyako Ohkubo, Koutarou Suzuki and Shingo Kinoshita, Cryptographic approach to "Privacy-Friendly" tags, RFID Privacy Workshop, MIT, MA, USA. November 2003.

[20] A. Juels, RFID security and privacy: A research survey, IEEE Journal on Selected Areas in Communications, 24(2), February 2006.

[21] G. Avoine, Security and privacy in RFID systems. http://lasecwww.epfl.ch/~gavoine/rfid/, 2006.
[22] Karsten Nohl and David Evans, Quantifying Information Leankage in Tree-Based Hash Protocols, ICICS 2006, LNCS 4307, pp228-237, 2006, Springer-Verlag Berlin Heidelberg 2006.